Perception of blending in stereo motion panoramas

Vincent Couture, Université de Montréal
Michael S. Langer, McGill University
Sébastien Roy, Université de Montréal

Most methods for synthesizing panoramas assume that the scene is static. A few methods have been proposed for synthesizing stereo or motion panoramas, but there has been little attempt to synthesize panoramas that have both stereo and motion. One faces several challenges in synthesizing stereo motion panoramas, for example, to ensure temporal synchronization between left and right views in each frame, to avoid spatial distortion of moving objects, and to continuously loop the video in time. We have recently developed a stereo motion panorama method that tries to address some of these challenges. The method blends space-time regions of a video XYT volume, such that the blending regions are distinct and translate over time. This paper presents a perception experiment that evaluates certain aspects of the method, namely how well observers can detect such blending regions. We measure detection time thresholds for different blending widths and for different scenes, and for monoscopic versus stereoscopic videos. Our results suggest that blending may be more effective in image regions that do not contain coherent moving objects that can be tracked over time. For example, we found moving water and partly transparent smoke were more effectively blended than swaying branches. We also found that performance in the task was roughly the same for mono versus stereo videos.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation—*Viewing algorithms*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual reality*

General Terms: Stereo, Perception

Additional Key Words and Phrases: visual perception, blending, motion, omnistereo

ACM Reference Format:

V. Couture, M.S. Langer and S. Roy. 2012. Perception of blending in stereo motion panoramas. ACM Trans. Appl. Percept. 2, 3, Article 1 (May 2010), 10 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Traditional stereo (3D) cinema uses two cameras with largely overlapping field of views. Well-known issues arise at the boundary of the field of view, namely 3D points that are rendered to be in front of the screen must not overlap the frame boundary, since this produces incorrect occlusion cues which are known as stereo window violations [Mendiburu 2009]. Correctly avoiding such window violations is one of several key technical challenges which must be addressed for stereo cinema to be successful.

Author's addresses: V. Couture and S. Roy, Université de Montréal, pavillon André-Aisenstadt (DIRO), 2920 chemin de la Tour, Montréal (QC), H3T 1J4; M.S. Langer, McGill University, McConnell Eng. room 318, 3480 University, Montreal (QC), Canada, H3A0E9.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

One way to reduce stereo window violations is to capture and display images with a very wide field of view, for example, a panorama. Many methods have been developed for synthesizing panoramas. For static scenes, these methods are now a standard part of the software toolkit of basic consumer level digital cameras [Woeste 2009]. However, these tools are for *monoscopic* images, that is, a set of images captured from a single viewpoint (not stereo images). Making stereo panoramas remains very challenging, even for static scenes. The problem of making panoramas with stereo and motion has, until recently, been so challenging that it has not been addressed at all.

Stereo motion panoramas could be applied in immersive 3D cinema. They could also be useful for standard field of view displays such as a (stereo) computer monitor. For example, a stereo motion panorama could allow a user to pan over a rendered 360 degree view. This would constitute a stereo motion extension of Quicktime or Google StreetView. Here we have in mind a case in which the motion is a texture such as waves on a river or lake, trees blowing in the wind, or a flag waving. For such motion textures, the video could loop over a few seconds [Schödl et al. 2000].

We have recently developed a method for synthesizing stereo-motion panoramas [Couture et al. 2011] which extends previous stereo panorama (also called *omnistereo*) methods that assume static scenes only. To our knowledge, our method is the only method that has been proposed for capturing panoramas that have stereo and motion. The purpose of the present paper is to examine conditions in which a method such as ours works well or not well. We present the results of an initial perceptual study.

The paper is organized as follows. Section 2 gives a brief overview of previous work on static stereo panoramas and on monoscopic motion panoramas. Section 3 gives a brief description of the method we have developed for stereo-motion panoramas [Couture et al. 2011]. We include this description of the method (rather than just citing it) in order to motivate some of the design choices in our perception experiment, in particular, the moving blending window. A reader who is more interested in the perception experiments may skip Sec. 3 and proceed directly to Sec. 4 where the perception experiment is described and results are presented. We conclude in Section 5.

2. PREVIOUS WORK

2.1 Static stereo panoramas

Existing computer vision methods for generating *static* stereo panoramas typically gather several small horizontal field of view images, namely vertical *slits*, from cameras rotating off-axis and then make a mosaic of these slits [Huang and Hung 1998; Ishiguro et al. 1992; Naemura et al. 1998; Peleg et al. 2001]. One of the reason for using slits is that, when the stereo rig is rotated, each camera rotates but also translates. This translation component causes parallax between images taken by same camera. If one were instead to stitch together a small number of regions of the left (or right) camera images that have large horizontal fields of view (i.e. not slits), then the parallax would produce visible seams at the region boundaries. For monoscopic panoramas, one solves this parallax problem by using a tripod mount that allows one to rotate the camera about its nodal point [Woeste 2009]. For stereo panoramas, this is not possible since cameras must translate as well as rotate in order to provide disparities in all directions around the observer [Seitz et al. 2002].

While slit-based stereo panorama methods work well for static scenes, they do not generalize well to dynamic scenes. If one were to apply a slit-based static omnistereo method to a dynamic scene, one would obtain left and right panoramas in which a moving object larger than a slit would be distorted spatially, because no slit is captured at the same time.

2.2 Monoscopic dynamic panoramas

A related set of methods have been developed for monoscopic panoramas with dynamic scenes. These methods consider a video as a space-time volume. One rearranges either slices [Rav-Acha et al. 2007] or small 3D blocks [Agarwala et al. 2005] from the volume, such that one avoids visual seams between the slices or blocks. The size and shape of blocks or slices are optimized using an intensity-based graph-cut minimization. In [Agarwala et al. 2005], a gradient blending method [Perez et al. 2003] is also performed to reduce the visibility of remaining seams between blocks while keeping the gradients inside each block.

The above methods have been used successfully to generate panoramas from videos taken by a (purely) rotating camera. To synthesize stereo motion videos, taking independent left and right spatial-temporal slices or small volumes would not in general lead to correct stereo disparities, because of scene motion. One might try to extend the panoramic video methods to stereo by enforcing a constraint on stereo motion consistency based on binocular disparities information (which would need to be estimated) but this would significantly increase the already large computational complexity of these methods.

In the next section we review our recently proposed method for stereo motion panoramas [Couture et al. 2011], which is based on blending at the boundaries of full frame stereo motion videos. Sec. 4 is motivated by the method, namely it examines how fast observers can detect the location of moving blended regions in a stereo motion video.

3. STEREO-MOTION PANORAMA METHOD

Like most stereo panorama methods, the method we presented in [Couture et al. 2011] uses a conventional stereo rig which is rotated around the mid-point between the two cameras, such that the two cameras each follow a circular motion. Each camera captures a space-time volume of images. The videos for a full rotation are about 2 minutes each, *i.e.* a few thousand stereo frames. What is new in our method is that our panoramas have both stereo and motion.

Let the raw left and right videos have N frames each. We assume that the camera turns 360 degrees, so frame N is where the camera completes the full turnaround, *i.e.* frame N is geometrically registered with frame 0. (This can be relaxed in practice, as only a partial turnaround can be performed.) Camera calibration and image registration yield left and right XYT volumes, composed of frames that shift over time as the cameras move. The image volumes are then resampled so that the rotational speed is uniform and the boundaries of the left and right image volumes are straight.

If we were to display the calibrated/registered image sequence in stereo on a cylindrical projection screen, one would see the scene through a window translating slowly over time, namely one would see the scene in stereo as captured by the rotating camera and projected in the correct direction. At any time, one would see only the field of view of the stereo camera. The problem that our method solves is to take this long stereo video as input and to synthesize a short looping panoramic stereo video which is defined over the entire cylinder. The main idea is to partition each of the stereo XYT volumes of the input video into parallelepiped blocks and to blend the blocks together to form left and right video textures.

The idea is illustrated in Fig. 1a,b. Suppose that, as the stereo rig rotates, it takes T frames for each visual direction in the scene to enter the field of view at the right edge and to exit the field of view at the left edge. We partition the entire image sequence into a set of consecutive blocks, each of T frames and then re-align the blocks so that all blocks start at the same frame ($T = 0$). In this example, the entire video is 120 seconds and is partitioned into five blocks that are 24 seconds each.

Consider a scene point that enters the field of view somewhere on the right diagonal of the first block. Since this point is visible for T frames, its path extends in time beyond frame T. When the blocks are aligned as in Fig. 1(b), the path followed by this point wraps around from frame $T - 1$ to frame 0. That is, the original video goes from frame $T - 1$ to frame T, and in the synthesized video frame T becomes frame 0. In general, frame i in the input video maps to frame $i \bmod T$ in the panoramic video.

As shown in [Couture et al. 2010], static scene points will have correctly aligned horizontal positions at the left and right boundaries (see Fig. 1(c)), but there will be small misalignments in vertical position for near points seen at the top or bottom of the field of view (discussed in [Couture et al. 2011]). These misalignments

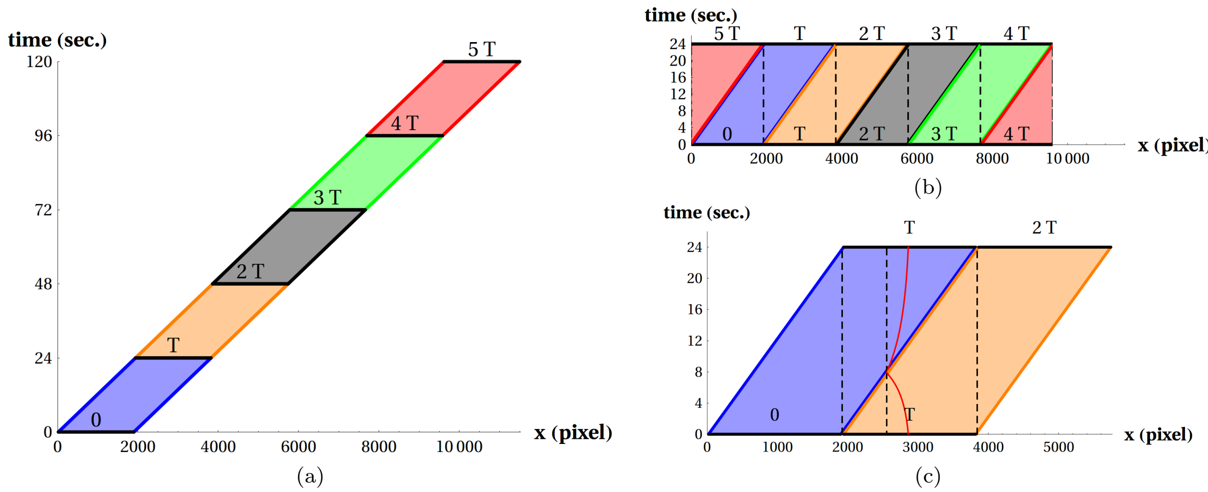


Fig. 1. For a frame sequence captured by a camera performing a full turnaround in $N = 5T$ seconds at constant speed: (a) the full original space-time volume divided in five non-overlapping blocks. (b) The blocks are aligned to start at the same time. (c) Motion paths of two static objects, one far away (central dashed vertical line) and one close-by (red curve). The curve in the latter is caused by motion parallax, that is, the cameras rotating off-axis. Despite using full frames, the motion paths are continuous and loopable.

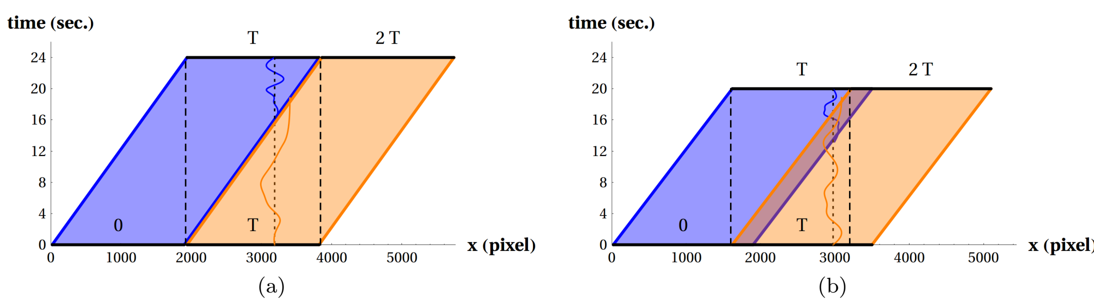


Fig. 2. (a) The motion of a visible scene point is continuous at the temporal boundary (horizontal edge), but there will be a motion discontinuity at the spatial boundary (seam), namely the diagonal edge in this figure. (b) For an object moving in time (a leaf, for instance), motion is blended over the overlap between the two blocks.

occur because the two images are captured from different positions. This is similar geometrically to the vertical disparities that can occur between left and right views.

For moving scene points, the misalignments in positions and velocities at the left and right boundaries may be more severe. Fig. 2(a)). To reduce the visibility in the seams where the boundaries meet, we overlap and blend the adjacent blocks near these boundaries. This requires decreasing the duration T of each block and shifting the block by the number of pixels covered during that decrease in duration. In Fig. 2(b), T has been decreased from 24 to 20 seconds. In the overlap region, we blend the frames together using a linear ramp function. Note that within the blending region, the stereo disparities of each moving object and its ghost are correct. As such, the depth cue of a moving object and its ghost do not conflict. Although blending leads to a duplication of moving points, the blended regions are distributed sparsely in the panoramic video, and their location changes over time. Moving scene points will therefore not be “ghosted” for most of the time. The perception experiment we describe in the next section examines how fast observers can detect the location of such blended regions.

Fig. 3. The five scenes used in the experiment, `flowers`, `bush`, `smoke`, `lake`, `river`.

We rendered example videos [Vision3D 2011] using the above method and displayed them as full panoramas on the cylindrical screen in our lab. Informally we observed that the stereo experience was excellent and that the blended regions were typically not noticeable, although the ghosting could be observed under scrutiny and became easier to notice once we knew what to look for and where to look in each scene. This was so, both for ourselves (authors) and for visitors to our lab.

4. PERCEPTION EXPERIMENTS

To study the detectability of blending in a more formal way, we carried out a perception experiment, The experiment used a small set of videos which contained a range of motions. We measured how fast observers could detect moving blended regions for different blending widths and for different scenes. We also compared performance with and without stereo, since it was not obvious whether stereo disparities would make the ghosting more or less salient.

4.1 Stimuli

Five new dynamic stereo scenes were captured using two parallel[1] cameras separated by 6.5 cm: `flowers, bush, smoke, lake,` and `river` (see Fig. 3). Unlike the scenes used in the method described above, these new scenes were shot in a fixed camera direction (no rotation). This was sufficient to investigate the visibility of the moving blending window and the ghosting that results.

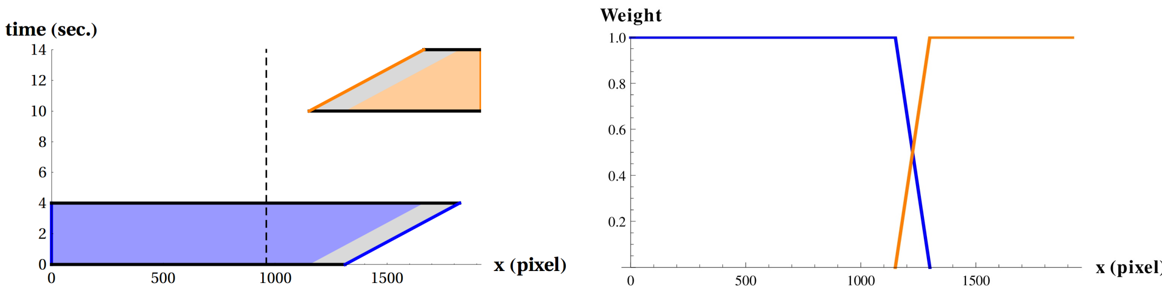
Two projectors were calibrated to produce a stereoscopic display on a large cylindrical screen, using the method of [Tardif et al. 2003]. Videos were displayed at 30 fps with a 16:9 projection aspect ratio and a resolution of 960×540 pixels. The viewing distance was 2.5 m which was the radius of the cylindrical screen. The field of view was about 45 degrees horizontal and 25 degrees vertical, which matched the field of view of the cameras. Observers wore stereo glasses for polarized viewing.

In each trial, a video was presented with a moving blending region (see Fig. 4). Each video was composed of two spatially overlapping blocks of the XYT volume (blue and orange, respectively, in Fig. 4) and captured several seconds apart. Thus, the spatial overlap was asynchronous in the blended region (gray in the figure), which produced ghosting of any moving objects. The blending region moved either left or right at a speed of about 3^o per second, which is similar to that of the panning cameras used in our stereo motion panorama method. The moving blending region in each trial was restricted either to the left or right half of the video.

The scenes were chosen to have similar left and right depth distributions. This was to avoid having subject's attention drawn consistently to one part of a scene, as happens when there is one dominant, close, spatially isolated and moving object, which is particularly salient in stereo.

Three conditions were examined. The first was the scene, namely we asked if ghosting was more visible in some of the videos than others. The second condition was the blending width. The third condition was mono versus stereo. In the mono version, the *same* image was displayed for both eyes, which is just the case of standard 2D cinema. In the stereo version, the images were shifted manually so that the screen disparity of closer objects was approximately 0. We did this to avoid objects appearing closer than the screen which can cause window violations[Mendiburu 2009]. Note that because of this shift the disparities did not correspond

[1]For an analysis of parallel versus converging cameras from a vision science perspective, see [Held and Banks 2008].



(a) (b)

Fig. 4. (a) To create a stimulus from a video, a blended region (gray) is created by overlapping the left part (blue) with the delayed right part (orange). The delay was set in this figure to 10 seconds for display purposes, but a delay between 25 and 35 seconds was used in practice. Here, the blended region is originally located at $\frac{3}{5}$ of the frame and moving right. Its width is $\frac{1}{12}$ the whole frame. (b) Blending function for the first frame of the stimulus.

exactly to what would have been experienced when looking at the real scene, i.e. our stereo display was non-orthostereoscopic.

Two versions of the experiment were performed:

—Version A tested each scene and 2 different blending widths, namely $\frac{1}{3}$ and $\frac{1}{12}$ of the frame width.

—Version B tested each scene for a single blending width, namely $\frac{1}{3}$ of the frame as in our panorama method, and compared stereo and monoscopic videos. By "monoscopic videos" here we just mean that the same video is shown to the left and right eye. This is just standard 2D video display.

Each of the two versions of the experiment had $10 = 5 \times 2$ conditions in total, namely 5 scenes $\times$ 2 independent conditions.

4.2 Observers

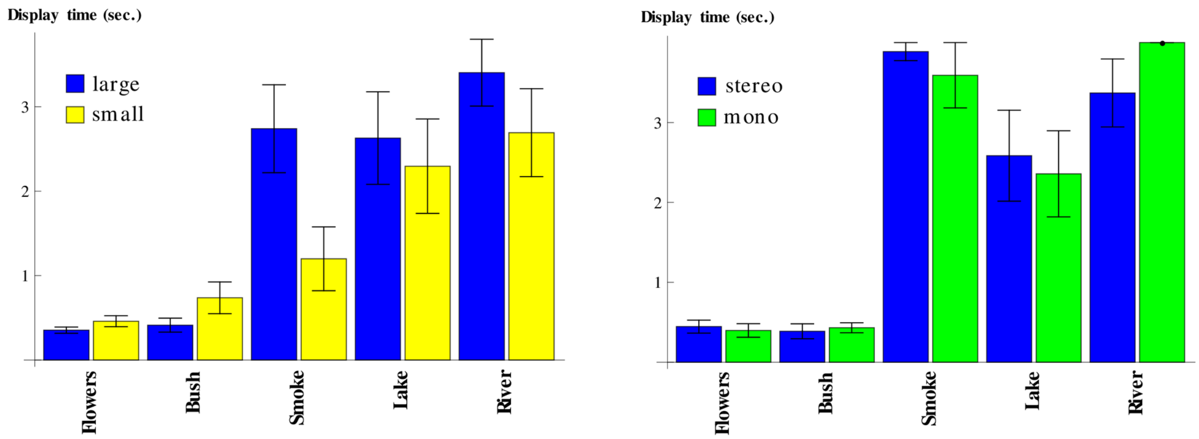
Twenty subjects participated. Ages ranged 18 to 65, and 17 were men and 3 women. Participants had varying degrees of familiarity with stereo cinema. All participants gave their formal consent based on a brief description about the motivation of the experiments.

Stereovision of each participant was tested using a random dot stereogram for which the disparities of two squares were adjusted so that one was to be perceived closer than the screen, and the other further away. The two squares were respectively located to the left and right of the stereogram and participants were asked to say which square was closer. This simple test was performed several times, each time randomly flipping the location of the squares. Two of the male participants failed the test and their results were not included in Fig. 5.

4.3 Procedure

For each condition, we wanted to know the minimum display time required to detect if the blended region was located in the left or right half. Each participant ran either version A or B of the experiment. The task in both versions was the same, namely to identify whether blending artifacts were present in the left or right half of the video. Before the experiment began, typical blending artifacts such as ghosting or reduced contrast were shown to the participants using the `flowers` and `lake` sequences. Observers were told that they needed to detect these artifacts. However, they were not told what caused these artifacts. In this sense, observers were naive.

The procedure for the experiment was as follows. Before each trial, a small cross appeared which indicated the screen center and thus defined the left and right halves. Observers were allowed to fixate whereever they



(a) (b)

Fig. 5. Detection time thresholds for (a) version A which compared large versus small blending widths, and (b) version B which compared mono vs. stereo presentations. The blue bars in the two experiments correspond to the same stimuli, namely a large blending width and stereo display.

wished and to move their heads. Although this was less controlled than what is typical in a psychophysics experiment, we felt that such a procedure would be better representative of how people might locate artifacts in a real stereo video such as in our rendered panoramic stereo videos. In each trial, the stimuli video was displayed for a varying duration (the independent variable). After the video finished, participants answered if the blended region was to the left or the right of the midpoint, using joystick buttons.

The 10 conditions were randomly interleaved. For each condition, the initial display time was 2 seconds. The display time was then updated to compute a detection threshold of about 75 % probability of being correct. [2] We used the up/down method proposed by [Dixon and Mood 1948], later transformed by [Wetherill and Levitt 1965]. Following the 1up/2down rule, the display time increased after every incorrect response, and decreased after two consecutive correct responses. This strategy made the display time increase if a participant guessed each response. The increase and decrease time factor was set to $\sqrt{2}$. Our termination criteria considered reversals of direction [Kingdom and Prins 2010]. A condition was stopped after 10 such reversals. To avoid losing time over a sequence too hard for a participant, we also stopped if a participant missed at a display time of 4 seconds. For faster convergence towards the threshold, we used the 1up/1down rule until the 1st reversal as suggested in [Wetherill and Levitt 1965]. Typically, each observer did about a few hundred trials.

For each condition, the threshold display time was computed as the average display time at the last 8 reversals. If a participant missed at a display time of 4 seconds, then this (maximum) time was considered as the threshold instead of the average.

4.4 Results

Mean thresholds across observers as well as the standard error of the mean are shown in Fig. 5. Our main finding is that, for both versions A and B of the experiment, the type of scene had a large effect. Thresholds were lower in the vegetation scenes (`flowers` and `bush`) than in the other scenes (`smoke`, `river` and `lake`), that is, blending was easier to detect in the vegetation scenes. Note that the clipping of the thresholds at

[2]We used this method rather than allowing for infinite display time (and computing reaction times) since in a pilot study we found that some observers took too long to respond and this led to high variabilities.

4 seconds occured for the `smoke`, `river` and `lake` scenes, and so their thresholds are larger and so their differences from the vegetation scenes are larger as well.

The other conditions had smaller effects, if any. For version A, large blending widths produced lower thresholds (suggesting the task was easier) in the vegetation scenes, but higher thresholds for the smoke and water scenes. For version B, mono and stereo conditions gave similar thresholds, which suggests that stereoscopic presentation did not make the task easier (at least for the scenes we tested). For further indirect support of this claim, we tested the two people that failed the stereo test. They gave similar thresholds for experiment A as the other participants (data not shown).

A final observation is that participants mentioned in the debriefing sessions that they had used different strategies to improve performance for very small display times. Some observers looked only on one side (left or right) and tried to detect if there was blending or not. If no blending was seen, then their chosen answer was the other side. Other observers reported learning specific scene locations and details where blending was more apparent. Since only five scenes were used, we expect that there were indeed learning effects with this experiment and that observers likely adopted strategies that allowed them to improve their performance as they saw more examples. However, since learning can occur for all stimulus conditions, we believe it does not explain the very large differences we have found between scene types, namely between the moving vegetation scenes versus the smoke and water scenes.

4.5 Discussion

The large differences in thresholds for the vegetation scenes versus the smoke and water scenes suggest that the blending method is much better suited for some types of scenes than others. Blending seems to work well for the smoke and water scenes in that it is quite difficult to detect even after several seconds of viewing and much practice. We believe that blending works well for these scenes because there are no well defined persistent features that the visual system can track across many frames. Smoke is partly transparent anyhow and so it is similar in appearance to alpha blending. When blending for the smoke scene was detectable, this seemed to be because the blending window was a vertically oriented ramp which is not a feature of natural smoke and hence appears unnatural. Note that the detection threshold was lower for the small blending width, since the vertical ramp had higher contrast and was more visible.

For the water scenes, features on the water surface appear and disappear rapidly over time as waves move. When blending is noticeable, it seems not to be because of ghosting of features. Rather it was because the image contrast was slightly lower in the blending region which is due to averaging. This is a different type of cue that observers may have learned during the experiment, though the data indicate that this cue had only limited benefit in the task. What is most interesting about the water scenes is how well they blended. Previous panoramic video methods which are based on graph cuts have used water as a compelling example of the success of their methods. While we do not dispute the success of these methods on such scenes, we would argue that such computationally intensive methods are overkill for water. Our data suggest that a much simpler technique namely blending is sufficient for building panoramas of these scenes.

The vegetation scenes are more problematic for a method based on blending. In the blended region, well-defined moving features such as leaves and swaying branches appear and disappear continuously through time, as if they had time-varying transparency. This appears unnatural for such features. Moreover, because the features have well defined shape and size which are preserved over time, the blending produces ghosting which is visually salient and attracts attention. In informal studies with other scenes, we have found the same results qualitatively. Well-defined objects that translate over time do not blend well. In order for blending to be effective, the motion needs to be more complex than a pure translation.

5. CONCLUSIONS AND FUTURE WORK

Blending is a common technique for combining images, either for creating mosaics or panoramas. Our perception experiment tested whether blending could be effective for combining two stereo motion videos. We found that detection of blending did not differ significantly between the stereo and mono conditions, at least for the scenes that we tested. We also found that blending was easier to detect for some scenes than others. In particular, when well-defined trackable moving objects such as branches or clusters of leaves are blended, visible ghosting occurs and can be detected. For scenes such as water flows or smoke for which motion is harder to track, blending was much more difficult to detect or not detected at all. This is somewhat surprising as panoramas of water scenes are usually considered to be challenging and have been used in the past to demonstrate the effectiveness of previous methods which are computationally expensive [Kwatra et al. 2003; Agarwala et al. 2005; Rav-Acha et al. 2007]. Our results suggest that a simple method namely blending is sufficient for water scenes.

Future work could investigate the perceptual effects of blending for different kinds of scene and in particular to measure the visibility of ghosting artifacts and various other distortions[Vangorp et al. 2011; Daly et al. 2011] as a function of scene category. Part of that analysis might require one to determine ranges of motion parameters in which different artifacts are visible. For example, for the vegetation scenes, how much motion is needed to produce visible ghosting? What are the amplitudes of the motion vectors and how coherent do the motions need to be? Similarly, for water scenes, how much of a decrease in contrast is needed for an observer to detect it? How does the threshold contrast depend on the masking which is due to the water motion?

Of course, we would also like to improve the blending method for the cases in which simple blending is not sufficient. For example, how can one use different blending functions at different spatial scales [Burt and Adelson 1983], or avoid blending or ghosting by finding an optimal seam between two image regions [Szeliski 2010] ? Recent advances in motion texture modelling may be applicable here [Derpanis and Wildes 2012], for example, one could recognize the type of motion automatically and then select a blending or video stitching method that is suitable for the type of motion (and stereo disparities) that are present there.

REFERENCES

Agarwala, A., Zheng, K., Pal, C., Agrawala, M., Cohen, M., Curless, B., Salesin, D., and Szeliski, R. 2005. Panoramic video textures. *ACM Transactions on Graphics 24,* 3, 821–827.

Burt, P. J. and Adelson, E. H. 1983. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics 2,* 4, 217–236.

Couture, V., Langer, M. S., and Roy, S. 2010. Capturing non-periodic omnistereo motions. In *10th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS).* Zaragoza, Spain.

Couture, V., Langer, M. S., and Roy, S. 2011. Panoramic stereo video textures. *IEEE International Conference on Computer Vision (ICCV)*, 1251–1258.

Daly, S., Held, R., and Hoffman, D. 2011. Perceptual issues in stereoscopic signal processing. *IEEE Transactions on Broadcasting 57,* 2 (June), 347 –361.

Derpanis, K. G. and Wildes, R. P. 2012. Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 34,* 6 (June), 1193–1205.

Dixon, W. and Mood, A. M. 1948. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association 43*, 109–126.

Held, R. T. and Banks, M. S. 2008. Misperceptions in stereoscopic displays: a vision science perspective. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization (APGV).* ACM, New York, NY, USA, 23–32.

Huang, H.-C. and Hung, Y.-P. 1998. Panoramic stereo imaging system with automatic disparity warping and seaming. *Graphical Models and Image Processing 60,* 3, 196–208.

Ishiguro, H., Yamamoto, M., and Tsuji, S. 1992. Omni-directional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14,* 2, 257–262.

Kingdom, F. A. and Prins, N. 2010. *Psychophysics: A Practical Introduction.* Academic Press.

KWATRA, V., SCHDL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics 22,* 3 (July), 277–286.

MENDIBURU, B. 2009. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen.* Focal Press.

NAEMURA, T., KANEKO, M., AND HARASHIMA, H. 1998. Multi-user immersive stereo. *IEEE International Conference on Image Processing 1*, 903.

PELEG, S., BEN-EZRA, M., AND PRITCH, Y. 2001. Omnistereo: Panoramic stereo imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23,* 3, 279–290.

PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. In *SIGGRAPH Conference Proceedings.* ACM Press, New York, NY, USA, 313–318.

RAV-ACHA, A., PRITCH, Y., LISCHINSKI, D., AND PELEG, S. 2007. Dynamosaicing: Mosaicing of dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29,* 10, 1789–1801.

SCHÖDL, A., SZELISKI, R., SALESIN, D., AND ESSA, I. 2000. Video textures. In *SIGGRAPH Conference Proceedings.* ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 489–498.

SEITZ, S., KALAI, A., AND SHUM, H. 2002. Omnivergent stereo. *International Journal of Computer Vision 48,* 3 (July), 159–172.

SZELISKI, R. 2010. *Computer Vision : Algorithms and Applications.* Springer-Verlag New York Inc. Chapter 9.3 (Compositing).

TARDIF, J.-P., ROY, S., AND TRUDEAU, M. 2003. Multi-projectors for arbitrary surfaces without explicit calibration nor reconstruction. *International Conference on 3-D Digital Imaging and Modeling*, 217–224.

VANGORP, P., CHAURASIA, G., LAFFONT, P.-Y., FLEMING, R., AND DRETTAKIS, G. 2011. Perception of visual artifacts in image-based rendering of faades. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering) 30,* 4 (July), 1241–1250.

VISION3D. 2011. `http://vision3d.iro.umontreal.ca/en/projects/omnistereo/`.

WETHERILL, G. B. AND LEVITT, H. 1965. Sequential estimation of points on a psychometric function. *The British journal of Mathematical and Statistical Psychology 18*, 1–10.

WOESTE, H. 2009. *Mastering Digital Panoramic Photography.* Rocky Nook.

Received Month YYYY; revised Month YYYY; accepted Month YYYY